

**EDUCATIONAL ACHIEVEMENT: EXPLANATIONS
AND IMPLICATIONS OF RECENT TRENDS**

The Congress of the United States
Congressional Budget Office

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, DC 20402



NOTES

Except where otherwise noted, dates used in this paper are school years rather than calendar years. For example, the results of a test administered in the fall of 1979 and the spring of 1980 are both labeled 1979. As a result, the dates used here are in some instances a year earlier than those in other published sources. This discrepancy is particularly common in the case of college admissions tests and other tests administered to high school seniors, which are often labeled in other sources in terms of the calendar year in which students would graduate.

Details in the text and tables of this report may not add to totals because of rounding.

PREFACE

At the request of the Subcommittee on Education, Arts, and the Humanities of the Senate Committee on Labor and Human Resources, the Congressional Budget Office (CBO) prepared this study of trends in the educational achievement of elementary and secondary school students. The first part of the study, *Trends in Educational Achievement*, published in April 1986, presented an analysis of the achievement trends themselves. This paper assesses the causes and implications of the trends. In accordance with CBO's mandate to provide objective and impartial analysis, neither volume contains recommendations.

Daniel Koretz of CBO's Human Resources Division prepared the analysis under the direction of Nancy M. Gordon and Martin D. Levine. Sherry Snyder edited the report. Ronald Moore typed and prepared the manuscript for publication.

Many individuals and organizations contributed in various ways to this report. Among the organizations that provided extensive data, some of which is unpublished, are the Iowa Testing Programs, the National Assessment of Educational Progress, and the A. C. Nielsen Company. Special thanks are due H. D. Hoover of the Iowa Testing Programs and Lawrence Rudner of the U. S. Department of Education, who provided insightful contributions at many stages of the project. Eric H. Hanushek of the University of Rochester, and Lyle V. Jones of the University of North Carolina, offered thorough and helpful reviews.

Edward M. Gramlich
Acting Director

August 1987



CONTENTS

SUMMARY

| | | |
|-----|------------------------------------------------------------------|----|
| I | INTRODUCTION | 1 |
| | The Context of the Current Controversy | 2 |
| | The Federal Role in Elementary and Secondary Education | 3 |
| | Recent Policy Initiatives | 7 |
| II | EDUCATIONAL ACHIEVEMENT: FACTS AND UNCERTAINTIES | 9 |
| | Test Scores as a Measure of Educational Achievement | 9 |
| | Patterns in the Achievement Data | 11 |
| | Important Gaps and Inconsistencies in the Achievement Data | 17 |
| III | APPROACHES TO EXPLAINING ACHIEVEMENT TRENDS | 21 |
| | Evaluating Evidence About Specific Factors | 22 |
| | Inferring Causes from General Patterns in the Test Score Data | 25 |
| IV | THE POSSIBLE CAUSES OF THE ACHIEVEMENT TRENDS | 27 |
| | Three Groups of Contributing Factors | 27 |
| | The Effects of Educational, Selection, and Societal Factors | 29 |
| V | IMPLICATIONS | 43 |
| | Assessing Educational Achievement | 43 |
| | Evaluating Educational Policies | 45 |
| | Improving Educational Achievement | 48 |

APPENDIX

53

| | |
|-----------------------------------------------------------|-----|
| Societal Factors | 53 |
| Changes in the Ethnic Composition of the Entire Cohort | 53 |
| Single-Parent Households | 54 |
| Family Size | 57 |
| Maternal Employment | 66 |
| Students' Use of Alcohol and Other Drugs | 68 |
| Television Viewing | 69 |
| Students' Attitudes and Motivation | 71 |
| Environmental Lead | 73 |
| Educational Factors | 76 |
| Teachers' Skills and Experience | 76 |
| State and Local Graduation Requirements | 81 |
| Coursework | 82 |
| Minimum-Competency Testing | 84 |
| Textbook Difficulty | 85 |
| Homework | 87 |
| Demands for Writing | 89 |
| Grade Inflation | 90 |
| Educational Programs for Disadvantaged Students | 91 |
| Desegregation | 95 |
| Selection Factors | 98 |
| Retention | 99 |
| Self-Selection | 102 |

TABLES

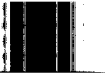
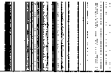
| | | |
|------|-------------------------------------------------------------------------------------------------------------------------|-----|
| 1. | The Effects of Various Educational, Societal, and Selection-Related Factors on Recent Trends in Test Scores | 32 |
| A-1. | Percent of High School Seniors Reporting Various Amounts of Time Spent on Homework Per Week, 1971 and 1979 School Years | 89 |
| A-2. | Trends in SAT-Verbal Scores and Reading Comprehension Among All High School Seniors, and SAT Candidates, 1959 to 1971 | 104 |

FIGURES

| | | |
|------|-------------------------------------------------------------------------------------------|-----|
| 1. | Shares of Elementary/Secondary Education Funding by Level of Government | 5 |
| 2. | Iowa Average Test Scores, Grades 5, 8, and 12, Differences from Post-1964 Low Point | 14 |
| A-1. | Average Total SAT and Average Birth Order (By year of birth) | 59 |
| A-2. | Television Viewing by Children, by Age (Average hours per week) | 71 |
| A-3. | Percent of 16- and 17-Year-Olds Enrolled Below College Level (All ethnic groups combined) | 101 |
| A-4. | Percent of Age Group in Modal Grade (Three-year moving averages) | 101 |

BOX

| | |
|----------------------------------------------|----|
| Have High School Test Scores Stopped Rising? | 19 |
|----------------------------------------------|----|



SUMMARY

The educational achievement of American elementary and secondary school students has been the focus of unusually intense scrutiny for several years. Strong public concern has been accompanied by extensive and continuing efforts at all levels of government to improve the public educational system.

Scores on standardized achievement tests have played a central role in this debate. Few issues were as critical to kindling the debate as was a growing public awareness that the test scores of American students declined markedly during the 1960s and 1970s and compared poorly with those of students in other countries. Many of the recent educational policy initiatives, such as stiffer standards for graduation from high school, were intended to counter these trends or to offset some of the factors (lax academic standards, in this case) that were presumed to have caused them. Moreover, many initiatives have increased the use of testing--of teachers as well as students--not only to measure achievement, but also to improve it.

Given the importance currently afforded scores on standardized tests, a careful appraisal of trends in test scores and their causes has significant implications for educational policy. *Trends in Educational Achievement*, a Congressional Budget Office study released in April 1986, assessed currently available data about trends in test scores and described some of the important limitations of standardized tests. This report analyzes possible causes of those trends and discusses implications for policy.

CURRENT INFORMATION ABOUT EDUCATIONAL ACHIEVEMENT

The existence of a sizable drop in test scores during the 1960s and 1970s has been well known for some time. The decline was remarkably pervasive, affecting many different types of students in most grades, in all regions of the United States, in Catholic as well as public schools, and even in Canadian schools. The drop was apparent in the results of different kinds of tests covering many subject areas. The deterioration was greater among

older students than in the early grades and affected higher-order skills such as reasoning and problem-solving more severely than more basic, rote skills.

The decline in scores was followed immediately by a widespread and significant rise. Perhaps because of the prominence of tests administered to senior-high students (for example, the Scholastic Aptitude Test, or SAT), many observers have mistakenly believed that the upturn did not start until the beginning of this decade (when SAT scores began to increase) and that it has been relatively inconsequential. Examination of a broader range of test data, however, shows that the upturn actually began by the mid-1970s and has been sizable. On certain tests administered to young children, for example, the upturn has more than overcome the previous decline.

Underlying the confusion about the timing of the upturn is a "cohort pattern" in the test scores that is central to understanding the possible causes of these trends. A cohort pattern is a change that affects children born in the same year, rather than children of various ages in school together in a given year (known as a "period effect"). The upturn typically began within a few years of the cohorts of children who were born in 1962 or 1963 and entered school in the late 1960s. The rise in scores first became apparent in the mid-1970s, when those children were in the middle elementary grades, and gradually moved into the higher grades as they progressed through school. Since then, successive cohorts of students have typically scored progressively higher. The lesser size and later onset of the rise in scores in the higher grades appears largely to reflect the smaller number of improving cohorts to have reached that level.

Several other variations in the trends are noteworthy. Black students and probably Hispanics have gained appreciably relative to their nonminority peers, although the gaps in scores between minority and nonminority groups remain large. The data also suggest that relative gains were made by students in schools with high minority enrollments and in disadvantaged urban communities.

Even though the recent rise in test scores has been substantial, the average level of performance on some tests remains well below what many educators would consider acceptable. Serious deficiencies can be found in all levels of skills, from the most rudimentary to the advanced. Moreover, many of these weaknesses will undoubtedly hinder students in their life outside of school. A disturbingly large proportion of American students, for example, are still unable to apply fundamental skills, such as simple mathematics, to situations encountered in everyday life.

GAPS IN CURRENT INFORMATION ABOUT EDUCATIONAL ACHIEVEMENT

Although considering a wide array of test data adds to the information that can be provided by one or a few tests, it also reveals a number of unanswered questions.

While some uncertainties simply reflect a scarcity of relevant data, others have arisen because existing tests--including those of high quality--sometimes provide inconsistent answers to even basic questions about educational achievement. For example, tests offer widely divergent estimates of the relative severity of the trends in different subject areas. Similarly, there are two recent nationally representative assessments of regional differences in achievement trends: one found particularly favorable trends in the South, while the other indicated a decline in the South that was comparable to or worse than that in other regions.

Another, potentially very important discrepancy among tests concerns the performance of the cohorts that have entered school in the last few years. While there is little reason to doubt that cohorts that have recently produced gains in the lower grades will continue to raise average scores as they progress through school, it is not clear whether incoming cohorts are continuing to outperform those that preceded them. Some tests show continuing gains in the lowest grades, while others suggest stagnation. Resolution of this question, which is important to any evaluation of the current wave of educational policy initiatives, will require information from additional tests administered over the next several years.

Such inconsistencies point to a critical, but widely ignored, limitation of standardized tests: even the best of current tests are only incomplete proxies for educational achievement. Most tests measure only some of the many skills required to master a broad subject area such as mathematics, for example. Consequently, the results of tests can differ from each other, often in ways that are unanticipated and difficult to explain. Moreover, important skills such as the ability to write well are difficult to assess using current standardized tests, and even data from several tests can yield inadequate information about them.

CAUSES OF THE ACHIEVEMENT TRENDS

Although a large number of diverse factors have been suggested as causes of the recent trends, many analysts are confident that one or a few factors can

account for much of the change shown by test scores over the past two decades. Moreover, many analysts believe that factors of a single type are responsible for those changes. The majority of them looks among educational factors for an explanation, while a smaller and less influential group expects the answer to be found in noneducational factors such as demographic trends and changes in students' use of alcohol and other drugs.

The available evidence, however, paints a much more complicated picture. The trends most likely resulted from the combined effects of numerous factors, both educational and noneducational. Moreover, to the extent that estimates are feasible, the individual contributions of those factors were typically modest. Two factors whose effects can be relatively well estimated and that appear to have made particularly substantial contributions to the trends--the changing ethnic composition of the school-age population and increasing family size--could each account for at most a fifth to a fourth of the total change in scores during portions of the achievement decline. The contributions of some other factors, while more difficult to estimate, appear to have been considerably smaller. Even taken together, the factors examined in this study provide only a partial explanation of the trends, and the limitations of the available data make it likely that any explanation will remain incomplete.

Perhaps because of the extensive attention paid to high school tests, many analysts who expect the achievement trends to have educational causes look to the late 1960s and 1970s--when the test scores of senior-high students were falling--for policies that might have caused the decline in scores. Similarly, many expect that the causes of the subsequent upturn can be found in the policies of the 1980s and perhaps the late 1970s.

While there is some truth in this view, it too is simpler than the data warrant. Some of the educational changes that contributed to the achievement trends were probably consistent in timing with trends in scores in the lower grades, not with scores at the senior-high level. The cohorts that produced the upturn in test scores entered school beginning in the late 1960s, and their improved performance was evident during their elementary school years. Thus, educational practices as early as the late 1960s and early 1970s--at least in elementary schools--might also have contributed to the rise in scores.

The factors that remain as plausible causes when systematic evidence is examined include a number of educational factors that often arise in the debate about achievement trends. A watering down of course content in secondary schools might have contributed to the decline in scores and might

help account for the greater severity of the decline in the higher grades. Changes in the amount of homework done by high school students, though relatively modest, might have contributed to both the decline and the subsequent upturn. Chapter 1 (the federally funded compensatory education program) could have contributed modestly to the relative gains of black and Hispanic students. Desegregation also might have contributed to the gains of blacks but apparently not to those of Hispanics, since the schools that Hispanics attend have become more segregated, not less.

The noneducational factors that could have contributed to the trends include some that are widely discussed and others that have received little notice in this context. Changes in family size that accompanied the baby boom and baby bust, which have received extensive attention, probably contributed moderately to both the decline and the upturn. Changes in the ethnic composition of the student body could account for perhaps a tenth to a fifth of the decline in test scores during the 1970s but probably impeded the rise in scores somewhat. Changes in students' use of alcohol and other drugs might have contributed to both the decline and the upturn and, like changes in coursework, might help explain the greater decline in the higher grades. A decrease in exposure to environmental lead--often discussed as an influence on children's health and cognitive functioning but rarely noted as a possible cause of trends in test scores--might have contributed in small measure to the upturn.

The list of factors that probably did not contribute significantly to the trends is more surprising, because it too includes factors that have gained widespread credence as possible causes. State graduation standards, for example, did not change significantly between 1974 and 1979 and therefore appear not to have contributed directly to the latter half of the achievement decline, and systematic data about requirements in earlier years are not available. Several commonly cited noneducational factors also do not weather close scrutiny. Whatever their effects on achievement in general, for example, neither television viewing nor the growing proportion of students living in single-parent households appear to have caused any significant share of the decline in test scores; the former did not change in ways that would have contributed to the trends in test scores, and the latter changed too little to have mattered in this context.

Finally, a number of commonly cited factors cannot be evaluated because existing data are inadequate. This gap in information is serious, because some of the factors that cannot be assessed have been important in the current debate and might have a substantial influence on test scores. These factors include local graduation requirements and students' motivation and attitudes toward education.

IMPLICATIONS

The analyses reported here have broad implications for assessing the condition of educational achievement and for formulating and evaluating educational policies.

Gauging the Condition of Educational Achievement

Because the currently available data leave important questions unanswered, additional national data from educational tests would clearly be helpful in assessing the achievement of American students.

The analysis in this report, however, argues strongly against relying solely on a single "national achievement test" for this additional information. A more reliable and informative, though costlier, alternative would be to maintain a number of tests, which ideally would vary in content and format. A comparison of several tests is often necessary to discern which results are consistent enough to provide a sound basis for policy, as evidenced by the several important instances in which the National Assessment of Educational Progress has yielded conclusions that are inconsistent with other data, and the wide variation in the results shown by other tests. Moreover, disparities in the results of different tests can themselves provide significant information. Because tests often stress different types of knowledge and skills, divergence in their results can reveal important facts about students' mastery of various aspects of a subject area.

For certain purposes, it would be critical to collect information about pertinent educational and noneducational factors, such as demographic trends and dropout rates, to accompany data from additional educational tests. Though costly to collect, such information would be important because the extent to which trends in test scores should be seen as real changes in students' achievement depends on the mix of factors responsible for them. At one extreme, trends in test scores attributable to educational factors, such as improved curricula, represent true changes in achievement. At the other extreme, trends in test scores that result from selection factors--that is, from changes in the selection of students to be tested--usually cannot be seen as actual changes in achievement. A drop in average test scores attributable to a decline in the dropout rate, for example, or to an increase in the number of less able students taking an optional college admissions test signifies nothing about the level of educational achievement of the school-age population as a whole. In between these two extremes are trends caused by societal factors--that is, noneducational factors other than selection changes. Such trends often would be seen as real changes in achievement, but their interpretation can vary depending on the factors involved and the question being addressed.

Evaluating Educational Policies

Trends in average test scores have become a common criterion for gauging the effectiveness of educational programs. The link between trends in test scores and educational policies, however, is far less straightforward than many people assume. Even when test data are sufficient to provide reliable information about students' achievement, they can lead to erroneous inferences about the effectiveness of educational programs.

Simple trends in test scores--that is, whether test scores are rising or falling--in themselves do not indicate whether policies are effective. Because many factors of different types (educational, societal, and selection-related) influence test scores, effective policies can be accompanied by falling scores, and rising scores can accompany policies that are actually detrimental. Accurate evaluation of a policy requires information on how trends have been deflected from the course they would have followed in the absence of that policy.

In the next few years, for example, simple trends in test scores will in many instances overestimate the effectiveness of educational policy initiatives because the current rise in scores antedates many of these policies and might well have continued in their absence, at least in the higher grades. In addition, the current emphasis on testing is likely to increase the extent to which teachers "teach to the test"--that is, tailor instruction specifically to raise scores. Regardless of whether increased teaching to the test is desirable, it is likely to make trends in test scores a distorted proxy for achievement.

In certain circumstances, however, simple trends in test scores will underestimate the effectiveness of educational initiatives. For example, scores may be depressed in districts undergoing unusually rapid demographic changes even if the policies carried out during that time are beneficial. Similarly, successful efforts to lower the dropout rate are likely to depress average scores.

Improving Educational Achievement

Many people have used trends in test scores and assumptions about their causes not only to formulate new educational policies, but also as a basis for presuming their effectiveness. Some assume that a few key factors that caused the decline of the 1960s and 1970s can be identified and that reversing those factors will cause scores to rise as markedly and as pervasively as they fell during those years.

Far from identifying a few key factors, however, this study suggests that changes in many, diverse educational factors might well be necessary to bring about increases in achievement as pervasive and large as the decline of the 1960s and 1970s. The individual contributions of educational factors to the recent trends were apparently modest. Moreover, since non-educational factors caused a sizable share of the change, even the effect of all educational causes combined, including factors not assessed here, fell substantially short of the total change in scores observed during those years. Thus, to bring about an increase as large and widespread as the decline would require a more powerful mix of educational changes than that which contributed to the decline.

This study thus suggests searching broadly for educational factors that might improve achievement. Focusing on factors that contributed to the trends of the recent past--for example, changes in the amount of homework assigned--might be productive. But the effects of those factors may be more modest than hoped, and limiting the search to them could exclude other factors of equal or greater importance. Factors whose contributions to the recent trends cannot be appraised for want of data, for example, include some--such as students' attitudes, demands for writing, and local graduation requirements--that might exert a powerful influence on students' learning. Even certain factors that apparently did not contribute to the recent trends--specifically, those, such as state graduation requirements, that did not change sufficiently during the relevant years--might also be important in the future.

Indeed, the results of this analysis suggest that the effectiveness of the current wave of initiatives should not be presumed on the basis of assumptions about what caused past trends. In many ways, the initiatives are more appropriately seen as an experiment than as a clear-cut response to the trends of the past two decades, and careful evaluation will be needed to assess their effects--both positive and negative.

Even though this study did not uncover the small number of key factors that many people would like to find, it does have several implications for the design of future initiatives. First, initiatives aimed primarily or entirely at the secondary level--for example, stiffened graduation requirements--even if beneficial, will miss an important part of the problem. The trends evident in the higher grades were also apparent in lower grades, and many of the skills in which deficiencies are particularly striking are taught in elementary and junior high schools.

Second, the data highlight the importance of improving higher-order skills, such as reasoning and problem-solving, at all grade levels. Even

though many rudimentary skills must be strengthened, policies that focus too much on rote skills and too little on reasoning and problem-solving will fail to address, and might even worsen, problems with higher-order skills that the test score data reveal to be particularly severe.

Finally, this analysis also suggests the need to focus on the performance of certain traditionally low-scoring groups but reaches no conclusions about the form that such initiatives should take. Although certain of these groups--for example, black students--have made appreciable gains, their level of achievement is still far below the national norm. The factors commonly advanced to account for these relative gains--desegregation and federally funded compensatory education--probably account for some of the improvement but leave much of it unexplained. Given the lack of an explanation for the rest of the improvement, there is a real danger that policies that were beneficial in this respect could be inadvertently discarded or undermined in the process of altering educational policy more generally. Only careful monitoring of the effects of the current wave of initiatives on the education of these students will clarify which of the changes further their recent gains and which erode them.

